

Review of Generalized Alternating Minimization Framework for Convergence of Incremental EM and Variational EM Algorithms

Huyen Nguyen¹

¹Department of Statistics, University of Connecticut, Storrs, CT, 06269

Abstract

The expectation-maximization (EM) algorithm is a well-known computational method to obtain the maximum likelihood estimator for the incomplete data problem. The algorithm consists of two steps: the first step is to compute the expectation of the likelihood function, i.e the E-step, and the second step is to maximize this expectation, i.e the M-step.

Keywords: expectation maximization, Monte Carlo methods, importance sampling, Robbin-Monro algorithm, non-linear mixed effects model

1 Introduction

Expectation Maximization (EM) is a popular algorithm to compute the maximum likelihood estimator in missing data problems. We assume that the complete data consists of (X, Z) with only X being observed. Let $\ell(\theta; X, Z)$ denote the complete-data log likelihood, in which θ is the unknown parameter vector that we wish to obtain the maximum likelihood estimate of. In other words, given an outcome of (X, Z) , we maximize the log likelihood function with respect to θ over its parameter space. The EM algorithm is an iterative method to compute the estimate.

- **E step:** In the E step, we compute the log likelihood of the complete data $\ell(\theta; X, Z)$ given the observed data X and current parameter estimate θ_t

$$(1) \quad Q(\theta|\theta_t) = E[\ell(\theta; X, Z)|X, \theta_t]$$

- **M step:** In the M step, we maximize $Q(\theta_{t+1}|\theta_t)$ over θ

$$(2) \quad \theta_{t+1} = \arg \max_{\theta} Q(\theta|\theta_t)$$

The maximum likelihood estimate of θ is then the best of local maximums obtained by iterations of the EM algorithm. To show why the EM algorithm works, we can use the Jensen's inequality by the concavity of the natural logarithm function (Roche [2011]). Let $p(\cdot|\cdot)$ denote the general conditional probability density function.

$$\begin{aligned} \ell(\theta; X) - \ell(\theta_t; X) &= \ln \frac{p(x|\theta)}{p(x|\theta_t)} \\ &= \ln \int \frac{p(x, z|\theta)}{p(x|\theta_t)} dz \\ (3) \quad &= \ln \int \frac{p(x, z|\theta)}{p(x, z|\theta_t)} p(z|x, \theta_t) dz \\ &\geq \int \ln \frac{p(x, z|\theta)}{p(x, z|\theta_t)} p(z|x, \theta_t) dz \\ &= Q(\theta|\theta_t) - Q(\theta_t|\theta_t) \end{aligned}$$

Thus, (3) shows that any value of θ that increases $Q(\theta|\theta_t)$ in (1) beyond $Q(\theta_t|\theta_t)$ will also increase $\ell(\theta; X)$ beyond $\ell(\theta_t; X)$ (Roche [2011]). There are numerous variants of the EM algorithm to adapt to practical applications. One of the issues commonly encountered is the intractability of the expectation in the E-step and an approximation method is needed to approximate the expectation.

In their paper, Gunawardana et al. [2005] introduced the generalized alternating minimization (GAM) framework and derive convergence results for variants of EM algorithm.

In section 2, we introduce the generalized alternating minimization framework introduced by this paper and their convergence theorem result. Section 3 and 4 present the results of GAM framework on the incremental EM Algorithm and the variational EM Algorithm, respectively.

2 Convergence Results of the GAM Framework

In this section, we introduce the generalized alternating minimization framework introduced by this paper and their convergence theorem result.

3 Incremental EM Algorithm

In this section, we present the results of GAM framework on the incremental EM Algorithm.

4 Variational EM Algorithm

In this section, we present the result of the GAM framework on the variational EM algorithm.

Appendix

This is the Appendix section.

References

- Gunawardana, A., Byrne, W., and Jordan, M. I. (2005). Convergence theorems for generalized alternating minimization procedures. *Journal of machine learning research*, 6(12).
- Roche, A. (2011). Em algorithm and variants: An informal tutorial. *arXiv preprint arXiv:1105.1476*.